Published in final edited form as:

Neurobiol Learn Mem. 2017 March ; 139: 179–188. doi:10.1016/j.nlm.2017.01.006.

Touch-screen visual reversal learning is mediated by value encoding and signal propagation in the orbitofrontal cortex

Kristin Marquardt¹, Rahul Sigdel¹, Jonathan L. Brigman^{1,2}

¹Department of Neurosciences, University of New Mexico School of Medicine, Albuquerque, NM

²New Mexico Alcohol Research Center, UNM Health Sciences Center, Albuquerque, NM

Abstract

Behavioral inflexibility is a common symptom of neuropsychiatric disorders which can have a major detrimental impact on quality of life. While the orbitofrontal cortex (OFC) has been strongly implicated in behavioral flexibility in rodents across paradigms, our understanding of how the OFC mediates these behaviors is rapidly adapting. Here we examined neuronal activity during reversal learning by coupling *in vivo* electrophysiological recording with a mouse touch-screen learning paradigm to further elucidate the role of the OFC in updating reward value. Single unit and oscillatory activity was recorded during well-learned discrimination and 3 distinct phases of reversal (early, chance and well-learned). During touch-screen performance, OFC neuronal firing tracked rewarded responses following a previous rewarded choice when behavior was well learned, but shifted to primarily track repeated errors following a previous error in early reversal. Spike activity tracked rewarded choices independent of previous trial outcome during chance reversal, and returned to the initial pattern of reward response at criterion. Analysis of spike coupling to oscillatory local field potentials showed that less frequently occurring behaviors had significantly fewer neurons locked to any oscillatory frequency. Together, these data support the role of the OFC in tracking the value of individual choices to inform future responses and suggest that oscillatory signaling may be involved in propagating responses to increase or decrease the likelihood that action is taken in the future. They further support the use of touch-screen paradigms in preclinical studies to more closely model clinical approaches to measuring behavioral flexibility.

Keywords

behavioral flexibility; *in vivo* electrophysiology; perseveration; spike firing; local field potential

Corresponding author: Jonathan L. Brigman, PhD, Department of Neurosciences, University of New Mexico School of Medicine, MSC08 4740, 1 University of New Mexico, Albuquerque, NM, USA 87131-0001, jbrigman@salud.unm.edu, Telephone: 505-272-2868, Fax: 505-272-8082.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Behavioral inflexibility is a common cognitive symptom of numerous neuropsychiatric and neurodevelopmental disorders, including but not limited to schizophrenia obsessive compulsive disorder, addiction, fetal alcohol- and autism spectrum- disorders. Inflexibility can have a profoundly negative impact on quality of life as a failure to adapt to changes in environmental conditions leads to intransigent patterns of behavior that affect relationships, financial management and the ability to maintain employment. The orbitofrontal cortex (OFC) has been implicated in mediating decision making and behavioral flexibility across species (Stalnaker et al., 2015). Targeted lesion and inactivation studies have demonstrated that OFC function is necessary for optimal reversal learning behavior, a hallmark task of behavioral flexibility (Hamilton and Brigman, 2015a, Izquierdo et al., 2016). Across modalities, reversal tasks require subjects to form expectations of outcome based on associative cues, perform based on those expectations and control changes in response to altered reward contingencies (Wilson et al., 2014, Costa et al., 2015, Jang et al., 2015, Saez et al., 2015, Stalnaker et al., 2015).

Studies using single and multi-unit recording during lever-press and spatial paradigms in the rat (Schoenbaum et al., 2000, Moorman and Aston-Jones, 2014) and olfactory/tactile paradigms in the mouse (Bissonette et al., 2008) suggest that the OFC responds with increased firing to an expected outcome, and tracks the efficient switching of reward value across reversal sessions. This supports the notion that OFC forms representations of expected outcomes based on previous trial outcomes, and that these representations are required to successfully switch choice behaviors when contingencies change (Padoa-Schioppa, 2007, Kennerley et al., 2011, Cai and Padoa-Schioppa, 2014). However, the OFC is also required to successfully monitor when learned actions fail to lead to the expected outcome. Responses to unexpected outcomes, or prediction errors, first described in midbrain dopamine neurons have been characterized in a subpopulation of OFC neurons which fire to an unexpected outcome (Thorpe et al., 1983, Mirenowicz and Schultz, 1994, Morris et al., 2006, Roesch et al., 2007). Inactivation of these neurons in the OFC can impair new learning once contingencies are changed, specifically when previous contingencies were well defined (Takahashi et al., 2009, Sul et al., 2010, Riceberg and Shapiro, 2012). Touch-screen reversal has similarly been shown to be sensitive to OFC lesion and targeted antagonism in mice, but to date, it has not been established that outcome value is similarly tracked in these more complex visual learning tasks in the rodent (Graybeal et al., 2011, Brigman et al., 2013).

In order to exert control over choice behaviors, changes in reward expectancies tracked by OFC must be communicated to downstream regions (Schoenbaum and Esber, 2010) involved in reward, habitual and goal directed behaviors (Schilman et al., 2008, van der Meer et al., 2010, Hoover and Vertes, 2011). It has been previously been shown that the dorsal striatum (dS) tracks behavioral responses during touch-screen learning, but how OFC value-signal propagation occurs has not been studied in this paradigm (Brigman et al., 2013). Given their hypothesized role in temporally coordinating neuronal firing within and between regions, oscillatory local field potentials (LFP) have been posited as the putative mechanism for propagating changes in neuronal firing across regions (Buzsaki and Draguhn,

2004, Womelsdorf et al., 2007, Cohen, 2014b). In addition, they have been proposed to behaviorally select single unit signals by frequency tuning (Fries et al. 2001). Recent evidence from rodent studies suggest that oscillations in the OFC can lock with spike-firing to distinct behaviors such as odor sampling and waiting for reward delivery (van Wingerden et al., 2010a, van Wingerden et al., 2010b) or spatial choice in a T-maze (Young and Shapiro, 2009). Understanding how local oscillations encode information and coordinate with single unit spikes during reward and error cues to signal changing reward contingencies could greatly improve our understanding of how OFC value encoding exerts influence over future choice behaviors across paradigms.

Touch-screen automated paradigms have become increasingly utilized to screen rodent models of numerous neuropsychiatric disorders (Marquardt et al., 2014, Yang et al., 2015, Copping et al., 2016, Leach et al., 2016) as these paradigms closely model tools used in the clinical assessment and may increase translational potential of preclinical studies (Mar et al., 2013, Talpos and Steckler, 2013, Hvoslef-Eide et al., 2016). While previous studies have demonstrated that lesion and/or inactivation of the region is sufficient to disrupt visual touch-screen reversal, it has not yet been demonstrated that the rodent OFC mediates reversal of complex visual stimuli in an analogous manner to those seen in primates using visual stimuli (Clarke et al., 2008), or to more species-specific stimuli such as spatial, olfactory or tactile stimuli in rodents (Hamilton and Brigman, 2015b).

Here we examined whether the OFC tracked reward expectancies during discrimination and reversal of visual stimuli in a touch-screen operant paradigm. We hypothesized that in agreement with lever, odor and spatial approaches, during distinct stages of touch-screen visual reversal, single units would signal changes in value and expectancy after choice behaviors differentially based on the previous trial response. We further hypothesized that these signals would be differentially coordinated with the local field potential to either increase or decrease likelihood that signals would be propagated to downstream regions to guide flexible behavior. To test this hypothesis, we utilized a touch-screen paradigm that provided immediate feedback on reward and error choices via concomitant tone and light cues and allows for recording of behavior and neuronal activity at well-established phases of reversal learning. Using this framework, we analyzed spike-firing and oscillatory activity during rewarded trials following a previously correct trial (*win-stay*) or following a previous error (*lose-shift*) and error trials that followed a rewarded trial (*regressive*) or followed another error in a series (*perseverative*) across discrimination and reversal to determine more precisely what behaviors the OFC encodes and potentially propagates at specific points of learning.

2. Materials and Methods

2.1 Subjects

Male C57BL/6J mice obtained from The Jackson Laboratory (Bar Harbor, ME) were housed in groupings of 2–4 per cage in a temperature- and humidity- controlled vivarium under a reverse 12 h light/dark cycle (lights off 0800 h). A total of 12 male mice were used for all experiments and tested during the dark phase. Beginning at 7 weeks of age, mice were food-restricted to 85% of their free-feeding body weight. Operant training began once mice

reached food-restricted weight. All experimental procedures were performed in accordance with the National Institutes of Health Guide for Care and Use of Laboratory Animals and were approved by the University of New Mexico Health Sciences Center Institutional Animal Care and Use Committee.

2.2 Touch-screen Apparatus

All operant behavior was conducted in a chamber measuring 21.6 x 17.8 x 12.7 cm (model # ENV-307W, Med Associates, St. Albans, VT), housed within a sound- and light- attenuating box (Med Associates, St. Albans, VT) as previously described (Marquardt et al., 2014). The standard grid floor of the chamber was covered with a solid acrylic plate to facilitate ambulation. A pellet dispenser delivering reward (14 mg dustless pellets; #F05684, BioServ, Frenchtown, NJ) into a magazine, a house-light, tone generator and an ultra-sensitive lever was located at one end of the chamber. At the opposite end of the chamber there was a touch-sensitive screen (Conclusive Solutions, U.K.) covered by a black acrylic aperture plate allowing two 2 x 5 cm touch areas separated by 0.5 cm and located at a height of 6.5 cm from the floor of the chamber. Stimulus presentation in the response windows and touches were controlled and recorded by the KLimbic Software Package (Conclusive Solutions, U.K.).

2.3 Pre-training

Mice were habituated to the operant chamber and to eating out of the pellet magazine by being placed in the chamber for up to 30 min with pellets available in the magazine. Mice retrieving 10 pellets within 30 min were moved onto pre-training. Mice began a three-stage pre-training regimen by first being trained to obtain reward by pressing a lever within the chamber on an FR1 schedule. Mice pressing and collecting 30 rewards in under 30 minutes were moved to touch training. During this stage, a lever press led to the initiation of a trial in which a white (variously-shaped) stimulus was presented in 1 of the 2 response windows. Lever press-initiation was included to clearly distinguish between initiation and reward seeking behaviors in later recording sessions. Throughout the paradigm images were spatially pseudorandomized preventing side bias and ensuring location was not an informative variable. The stimulus remained on the screen until a response was made. Touches in the blank response window had no effect, while a touch to the white stimulus resulted in reward delivery, immediately cued by a tone and illumination of the magazine light on the opposite side of the operant chamber from the touch screen. Mice initiating, touching and retrieving 30 pellets within 30 min were moved to the final stage of pre-training. This stage was identical to touch-training except that responses at the blank window during stimulus presentation produced an immediate 10 sec timeout, signaled by illumination of the house light, to discourage indiscriminate screen responding. Errors on this, and all subsequent stages, were followed by correction trials in which the same stimuli and left/right position was presented until a correct response was made. Mice making 75% (excluding correction trials) of their responses at a stimulus-containing window over a 30-trial session were moved onto discrimination.

2.4 Stereotaxic Array Implantation

After completing pre-training and at least two consecutive days of free-feeding, mice were anesthetized with isoflurane and placed in a stereotaxic alignment system (Kopf Instruments, Tujunga, CA) for implantation of a microelectrode array. The array (Innovative Neurophysiology, Durham, NC) comprised 16 individual 35 μm -diameter tungsten microelectrodes arranged into 2 bundles of 2x4 electrodes (150 μm row/column spacing, 2.75 mm spacing between bundles) targeting bilateral orbitofrontal cortex (center of array: AP +2.60, ML $\pm$ 1.38, DV -2.60). After 7 days of recovery, body weight reduction resumed and mice were given a post-surgery reminder consisting of the last pre-training session to ensure retention of pre-training criterion.

2.5 Discrimination and Reversal Learning

Following array implantation, all mice were tested on a pairwise discrimination reversal paradigm as previously described (Brigman et al., 2010b). Mice were first trained to discriminate two novel, approximately equiluminant stimuli (Fan/Marble (Brigman et al., 2006, Brigman et al., 2010a, Brigman et al., 2013, Marquardt et al., 2014)), presented spatially pseudorandomized across 30-first presentation trials (not including correction trials) per daily session (5 sec ITI) (Figure 1A). As in pre-training, responses at the correct stimulus resulted in reward, immediately cued by the onset of a 1 sec tone; responses at the incorrect stimulus resulted in timeout, immediately cued by a 10 sec house-light followed by correction trials until a correct response was made (Figure 1B). Correct stimuli were balanced across mice. Discrimination criterion was of 85% correct responding (excluding correction trials) over two consecutive sessions. Reversal training began on the session after discrimination criterion was attained. Here, the designation of correct versus incorrect stimuli was reversed for each mouse. Mice were trained on 30-trial daily sessions (same as for discrimination) to a criterion of 85% correct responding (excluding correction trials) over two consecutive sessions. In order to measure performance differences across distinct sessions, percent correct responses, total errors, reaction time (time from lever press initiation to screen touch) and magazine latency (time from screen touch to reward retrieval) were analyzed. In order to analyze use of feedback for learning, correct and incorrect responses were further categorized based on previous trial outcome: correct responses were characterized as *win-stay* (following correct response) or *lose-shift* (following an error trial), while error trials were characterized as *perseverative* (following an error trial) or *regressive* (following a correct response; Figure 1C).

2.6 Neurophysiological Recording

Neuronal activity was continuously recorded using a multichannel acquisition processor (OmniPlex, Plexon, Dallas, TX) as previously described (Brigman et al., 2013, DePoy et al., 2013). Single and oscillatory activity was captured during the following 4 stages: **Discrimination Criterion** = session of discrimination criterion attainment (% correct >85%), **Early Reversal** = first session of reversal where perseverative responding is highest (% correct <20%), **Chance Reversal** (Chance) = session of reversal where chance performance was re-attained (% correct =50%, and **Reversal Criterion**= session of reversal where criterion is re-attained (% correct >85%; Figure 1D). Continuous spike signal was

sampled at 40 kHz and waveforms were manually sorted during recording, based on manually set voltage threshold. Local field potential was sampled from the same electrodes at 1 kHz and automatically low band pass filtered at 200 Hz. Neuronal recording data was timestamped by responses from k-limbic software by TTL pulse to reward tone and punishment house light. At the completion of testing, array placement was verified via electrolytic lesions made by passing 100 μ A through the electrodes for 20 sec (S48 Square Pulse Stimulator, Grass Technologies, West Warwick, RI). Brains were removed post perfusion with 4% paraformaldehyde, 50 μ m coronal sections cut with a vibratome (Classic 1000 model, Vibratome, Bannockburn, IL), stained with cresyl violet and placement verified with reference to a mouse brain atlas (Paxinos and Franklin, 2001)(Figure 1E).

2.7 Waveforms Analysis

Waveforms were re-sorted offline using principal component analysis of spike clusters and visual inspection of waveform and inter-spike interval <1% shorter than 2ms using Offline Sorter (Plexon Inc, Dallas, Texas). Recording across multiple sessions increases the likelihood of repeatedly sampling individual units. However, tracking of individual units across sessions could not be verified, sorted putative neurons were therefore treated as independent units between sessions. TTL pulse timestamps recorded concurrently with neuronal data were used to create epochs of firing rate spanning 1 sec pre-choice to 3 sec post-choice in averaged bins of .05 sec with NeuroExplorer software (NeuroExplorer; NEX Technologies, Littleton, MA). Firing rate was analyzed during epochs of choice behaviors which were defined by the previous trial completed: correct responses following a previously correct response (*win-stay*) or a previous error (*lose-shift*), and errors following a previous error (*perseverative*) or correct trial (*regressive*). The 3 sec post-choice analysis window overlapped with immediate tone, during correct choice, and first 3 sec of house-light during incorrect trials allowing for analysis of immediate response to secondary associative cues. If reward retrieval occurred prior to the end of the 3 sec analysis window on correct trials, the epoch for that trial was truncated as to prevent overlap with reward signaling. Less than 5% of neurons showed baseline firing rate of >15Hz, were categorized as fast-spiking interneurons and excluded from analysis as previously described (Brigman et al., 2013). Normalized firing rates were calculated by Z-score of 3 sec post-choice to 1 sec pre-choice baseline. Pattern of firing rate changes were examined for the period during immediate cue delivery (choice $\rightarrow$ 1 sec post) and magazine approach (2 $\rightarrow$ 3 sec.) using repeated-measures ANOVA followed by the Newman-Keuls post-hoc test. The threshold for statistical significance was $P < 0.05$.

In addition to spike-firing activity changes, the proportion of the population of neurons that significantly increased their firing rate in the post-choice period (compared to individual pre-choice baseline using Student's t-test) were analyzed across type and discrimination reversal session via chi square. All choice- responsive neurons were analyzed, independent of significant response to other timestamped events as single unit signaling in the OFC is extremely heterogeneous and selection of single event- responsive neurons would unfairly represent overall response (Thorpe et al., 1983, Moorman and Aston-Jones, 2014, McMurray et al., 2016).

2.8 Spike Field Coupling

In order to compare timing of spikes relative to shifts in phase within the LFP analysis of spike field coupling was conducted (Cohen, 2014b). For each of the four defined trial types, single instances of spike firing across waveforms were marked and 500ms of oscillatory LFP data around each spike was set into an individual epoch with the spike occurrence at time point 0. The spike-LFP phase angle was computed as previously described (Cohen, 2014a) and used to calculate the paired phase consistency (PPC0) value (see Supplementary Material; (Vinck et al., 2010)). To control for unequal trial numbers across all behavioral bins and trial types 10 spike-locked LFPs were randomly selected to perform the PPC calculation over 1000 permutations to give an average, unbiased by spike number. Bins were combined for a final average of the 3 sec post-choice time epoch and PPC within each trial type, session and time epoch was compared to scrambled data within the same type. Significance was determined by a difference greater than two times the standard deviation from scrambled data.

3. Results

3.1 Behavior Profile of Touchscreen Reversal Learning

Mice with multichannel electrode arrays readily re-attained pre-training criterion (1.45 ± 0.4 sessions) and demonstrated a clear pattern of flexible shifting from a well-learned behavior to a new response. Mice progressed through stages of discrimination and reversal learning requiring similar numbers of trials as seen previously (Figure 2A; ANOVA effect of stage on correct trials: $F_{3,39}=72.43$, $P<.01$, followed by *post hoc* tests). Four distinct target trial types resulted from sorting based on the N-1 previous trial response, which allowed for the analysis of the microstructure of rewarded behavior. *Win-stay* trials (Figure 2B), were significantly more prominent on discrimination criterion, indicating a well-learned and beneficial response strategy. *Win-stay* response significantly decreased during early reversal before increasing in a step-wise pattern across chance reversal and returned to high levels upon re-attainment of criterion on reversal ($F_{3,10}=263.618$, $P<.001$ followed by *post-hoc* test). In contrast, *lose-shift* trials represent a positive change in strategy to obtain a reward due to the previous error response (Figure 2C). Low levels of this exploratory behavior are seen during criterion stages and early reversal. This behavior increases during chance reversal during which the new association is being learned, before again decreasing at reversal criterion ($F_{3,36}=18.117$, $P<.001$ followed by *post hoc* tests).

Analysis of total error choices as mice progressed through discrimination and reversal were similar to those seen in previous *in vivo* recording experiments (Figure 2D; ANOVA stage effect: $F_{3,39}=42.94$, $P<.01$, followed by *post hoc* tests). Analysis of the microstructure of error responses confirmed previous findings that *perseverative* (error-error; Figure 2E) responses significantly increase during early reversal, dominating all other responses before tapering off by chance and becoming virtually non-existent by criterion reversal ($F_{3,36}=26.359$, $P<.001$ followed by *post hoc* tests;). *Perseverative* errors continued during chance reversal, even with the presence of the more beneficial *win-stay* trials within the same session, indicating the difficulty in initiating a change in response strategy. Similar to *lose-shift* responses, *regressive* errors, a non-beneficial change in behavior, which is not

prompted by changing contingencies (reward-error; Figure 2F), occur at very low levels during well-learned criterion discrimination and increase during early reversal, but not significantly. *Regressive* errors significantly increase in conjunction with *lose-shift* trials as new associations are learned during chance reversal, before decreasing again when criterion reversal is obtained ($F_{3,36}=20.088$, $P<.001$ followed by *post hoc* tests;).

Analysis of secondary behavioral measures showed a small but significant increase in latency to make a screen response on early reversal (Figure S1B; $F_{3,11}=3.466$, $P<.05$). This is unlikely to be a change in general motivation as the latency to retrieve a reward showed no significant differences (Figure S1A), but an alteration in behavior caused by the shift in paradigm rules during early reversal.

3.2 Choice-Responsive Neurons are Learning Session Specific

We recorded single unit activity from 166 putative neurons. There was no significant difference in 1 sec pre-choice baseline between any trial types within a session, or between sessions (Figure S2). During distinct phases of discrimination and reversal populations of OFC neurons increased firing significantly to specific reward and error types. *Win-stay trials* were the dominant signaling type during well learned behavior, showing significantly more choice-responsive neuronal recruitment ($\chi^2=6.44$, $P<.05$; Figure 3A *right*). Analysis of all neurons found a significant increase in firing to both *win-stay* and *lose-shift* rewarded responses 2 seconds after correct choice response during reward approach (Figure 3A *left*; MAIN EFFECT OF TIME: $F_{3,291}=4.39$, $P<.001$). There was also a significant interaction between rewarded trial types, revealing that *win-stay* had a significantly increased firing rate compared to *lose-shift* trials (INTERACTION: $F_{3,291}=2.70$, $P<.05$). In contrast, there was a significant sustained increase in firing after both *regressive* and *perseverative* errors immediately after negative reinforcement cue onset during discrimination criterion (Figure 3A *center*; MAIN EFFECT OF TIME: $F_{3,210}=4.79$, $P<.001$). However, there was no significant difference in the strength or pattern of firing between error trial types.

During early reversal the percent of *win-stay* trial responsive neurons significantly decreased while *lose-shift* responsive neurons became the dominate rewarded trial type (Figure 3B *right*; $\chi^2=11.83$, $P<.01$). Neither *win-stay* nor *lose-shift* rewarded trials responded with a significant increase in firing rate (Figure 3B *left*). Both error trial types had more responsive neurons than discrimination criterion, with *perseverative* trial responsive neurons dominating ($\chi^2=22.78$, $P<.01$). *Perseverative* and *regressive* error trials led to an immediate and sustained significant increase in firing rate during early reversal (Figure 3B *center*; $F_{3,297}=4.79$, $P<.001$), with no significant difference between firing intensity.

During chance reversal, there continued to be more choice-responsive neurons to *lose-shift* rewarded trials than *win-stay* trials, but not significantly. However, the proportion of choice-responsive neurons to error trials overall decreased, and the proportion of *regressive* choice-responsive neurons are significantly higher than *perseverative* (Figure 3C *right*; $\chi^2=3.97$, $P<.05$). Neurons responsive to correct responses did increase firing rate to the end of the associative signal (tone), but did not reach significance. Similarly, neither error trial response led to an increase in firing.

When reversal criterion was attained, the pattern of choice-responsive neurons mirrored discrimination criterion. *Win-stay* responsive neurons were the dominant type of reward responsive neurons with no neurons responsive to *lose-shift* ($\chi^2=103.15$, $P<.01$). While slightly elevated over discrimination, there was no significant difference between proportion of neurons responsive to *regressive* or *perseverative* errors (Figure 3D *right*). As in discrimination criterion, there was a significant increase in firing to both rewarded responses two seconds after choice (Figure 3D *left*; MAIN EFFECT OF TIME: $F_{3,120}=5.73$, $P<.001$). Similarly, *win-stay* responses had a significantly increased firing rate compared to *lose-shift* trials (INTERACTION: $F_{3,120}=3.18$, $P<.05$). In contrast to discrimination, during reversal criterion, *perseverative* nor *regressive* error trials were followed by significant increases in firing during any epoch.

3.3 Spike-Field Coupling Tracks Reinforced Behavior

Comparisons of time-frequency power spectra did not reveal any consistent changes that were greater than two standard deviations from chance, during any session of reversal in any trial type (Figure S3A–E) indicating the time-frequency power spectra of each trial type was independent of learning session. However, analysis of synchrony between spike firing and oscillatory activity revealed distinctive patterns of coupling across learning sessions. Trials in which a switch in behavior occurred (*lose-shift* and *regressive*) had significantly greater than chance spike- theta field coupling (4–10 Hz) during discrimination criterion (Figure 4A). In contrast, both consistent behavioral trial types (*win-stay* and *perseverative* responses) had extensive cross- frequency decreases in synchrony that were two standard deviations *lower* than chance indicating a decrease in signal coordination. However, *win-stay* decreases were limited to beta (10–30 Hz) and portions of gamma (30–40 Hz), while neurons responsive to *perseverative* trials were de-coupled across every measured frequency (5–40 Hz).

Lose-shift, *regressive*, and *perseverative* responses returned to chance level of spike-field coupling during early reversal (Figure 4B). In contrast, *win-stay* trials were strongly de-synchronized (greater than two standard deviations less than chance) across every frequency tested, except for a small band between 11 and 15 Hz, mirroring decreases in signal coordination seen during perseverative trials on the previous learning stage. *Win-stay* trials return to chance levels of synchrony by chance reversal (Figure 4C *top*) while both error responses continue to have non-significant changes in spike field coupling (Figure 4C *bottom*). Only *lose-shift* trials had significant changes exceeding two standard deviations from chance in spike-field coupling during chance reversal, a small increase from 9–11 Hz and decrease from 15 to 25 Hz.

Patterns of spike field synchrony do not return to discrimination criterion patterns during reversal criterion. *Win-stay* trials at this stage were consistent with chance reversal with no significant changes in spike field coupling. This is in stark contrast to both *lose-shift* and *regressive* errors that became strongly decoupled across multiple frequency ranges (greater than two standard deviations lower than chance). *Lose-shift* responses were decoupled across all three-frequency bands of interest, while *regressive* is limited to changes in beta

and gamma frequencies. *Perseverative* PPC was immeasurable due to extremely low spike responsiveness, indicating it is strongly decoupled during criterion reversal.

4. Discussion

Consistent with previous studies in primate visual and rat spatial paradigms, our data suggest that the OFC distinctly encodes values of specific choices during different stages of learning in touch-screen visual reversal (Moorman and Aston-Jones, 2014, Rich and Wallis, 2016). We also saw dynamic alterations in coupling between neuronal spike-firing and oscillatory activity that suggests selective propagation of choice-responsive signals to influence future behaviors. Taken together, our data support the role of choice-responsive OFC neurons in encoding value in a rodent touch-screen task, and further suggests spike coupling with local oscillations as a putative mechanism by which signals are propagated to downstream regions to increase or decrease the likelihood of a behavioral action.

4.1 Spike-firing encodes value-expectancy across discrete behavioral choices

During touch-screen visual performance, OFC neurons showed distinct choice-responsive firing rates based on the likelihood a given response would lead to reward at that stage, suggesting that OFC choice-responsive neurons encode a pre-reward value expectancy (Young and Shapiro, 2009, Sul et al., 2010, Cai and Padoa-Schioppa, 2014, Moorman and Aston-Jones, 2014). During criterion performance, *win-stay* choice firing dominated all other signaling. Thus, this signal, which was increased when the expected value of the choice matched the outcome, is highly congruent with reward expectancy signaling in the OFC seen across rat spatial tasks and primate visual studies (Young and Shapiro, 2009, Sul et al., 2010, Cai and Padoa-Schioppa, 2014, Moorman and Aston-Jones, 2014). In contrast, during early reversal where outcome *least* matched the learned expectancy, firing shifted to robustly track error choices, suggesting these signals may reflect a negative expectancy violation. In non-human primates, spike-firing responses switch target responding within several trials (Thorpe et al., 1983). In the current study, where reversal takes several sessions, we see a clear response to the unexpected negative outcome on *perseverative* choice trials due to the unsignaled shift in stimuli contingencies. This suggests that during early reversal in our paradigm, OFC neuronal signaling that was previously tied to an expected reward shifts, to now signal the failure of that reward to occur. While we are unable to differentiate whether our recorded neuronal populations are consistent across sessions, our data suggest the OFC tracks choice values differentially when expectancies are consistently met (criterion) versus violated (early reversal). As we and others have shown loss of OFC function significantly increases these *perseverative* responses and extends the timeframe to exit the perseverative period (Clarke et al., 2008, Graybeal et al., 2011), we hypothesize this signaling is critically important to alter the choice-value has changed and facilitate flexible behavior. Similarly, *regressive* choice errors increased both in total number of responsive neurons and firing strength during this period. While both error types do not result in reward when it is expected, differences in signaling pattern, and number of responsive neurons in *regressive* compared to *perseverative* errors, indicate that there may be differences in downstream influences. It has been recently reported that the OFC does not differentially signal responses between stay and exploratory trials during lever press probabilistic reversal

in an extended post-choice analysis window (Amodeo et al., 2016). Our results suggest that differences in signaling between consistent versus exploratory trials may only occur immediately after the choice event and may not be captured when data is analyzed over extended periods. While firing rates did not change, we also detected a significant increase in number of choice-responsive neurons to *lose-shift* during early reversal, suggesting the OFC also tracks these infrequent unexpected rewards, as has been previously shown in a Pavlovian odor task in the rat (Takahashi et al., 2009)

The recruitment of choice-responsive neurons when expectations of value become ambiguous also provides evidence of online value encoding in the OFC. During chance reversal, when the animal is no longer highly perseverative, but has not learned the new association, *lose-shift* and *regressive* trials recruit more choice-responsive neurons than other responses. This switch in OFC recruitment may represent a shift to a more exploratory behavior pattern, as mice sample the outcome of different choices. While at criterion stages we found OFC firing strongly increased during reward approach, at chance reversal, firing increased to the immediate reward cue as described in non-human primates (Padoa-Schioppa and Assad, 2006). While our previous studies show that the OFC is not functionally necessary to establish the new choice values during touch-screen reversal (Brigman et al., 2013), the shift in OFC firing to immediate reward cues suggest the OFC tracks cue value to provide immediate feedback when the optimal choice is ambiguous. In contrast, when choice-values are well-learned, the cue may be too far removed in time from reward retrieval to continue this association, and therefore the approach and other un-intended secondary cues, like the pellet delivery, become informative for value.

4.2 Spike-field coupling correlates with behavioral strategy continuation or cessation

As reported previously in spatial reversal tasks, we did not detect large pattern changes in power across discrimination-reversal sessions (Young and Shapiro, 2011). In line with human EEG studies of consistent resting state power over time, analysis of power by trial types found distinct patterns that were consistent across sessions, suggesting OFC power patterns are set for particular behavioral responses (Porjesz et al., 2002). Local field potential power is a measure of the strength in the local oscillations, and changes in power can represent changes in activity or attention directed to the task suggesting that attentional capacity did not significantly vary across task stages (Buzsaki and Draguhn, 2004, Buzsaki et al., 2012, Calderone et al., 2014). The lack of changes in power across reversal suggests that meaningful value signals likely depend on dynamic coupling of varying event-dependent spike-firing events to the more consistent oscillatory signals.

Synchronization of spikes by local oscillations leads to coordinated large scale networks by amplifying or attenuating spike signals in multiple regions across the brain during spatial reversal and visual attention tasks (Womelsdorf et al., 2007, Gregoriou et al., 2009, Cohen, 2014b, Womelsdorf et al., 2014). Here we show that during reversal, the dynamics of the spike-field coupling and decoupling correlated with the stereotypical reversal behavioral pattern, suggesting amplification and attenuation of specific trials. When discrimination was well learned, theta spike-field coupling was increased during more exploratory choice-trials independent upon rewarded outcome. This is in contrast to an odor discrimination

task where theta spike-field coupling was linked to reward anticipation (van Wingerden et al., 2010a). During criterion performance, these exploratory trials occur infrequently, suggesting feedback generated by preferentially aligning signaling during these trial types to the oscillations may serve as a general monitoring function, for matching expectancy and outcome. Previous reports suggest spike desynchronization in low frequency ranges reduces spike co-occurrences focusing visual attention (Fries et al., 2001). Multi-frequency spike-field de-synchronization in the current visual reversal task may strongly impact behavior by decreasing the signaling influence of particular trial types. During well-learned behavior the OFC strongly decreases the value- signaling impact of any *perseverative* trials, thus promoting *win-stay* behavioral response. Early reversal, dominated by negative- value signaling in the OFC, responses to infrequent strings of *correct* choices are not propagated resulting in strings of errors.

Spike-field coupling differences between criterion stages likely reflect differences in the two stages from a value encoding perspective. During discrimination, stimulus choices have a fixed never varied value, while after reversal stimuli have multiple values, which vary based on *when* they are advantageous. Therefore, differences in spike-field coupling may be the result of reversal experience. These alterations may facilitate future reversals by holding both new and previous expectancies online, which would prime for future changes in reward associations (Boulougouris et al., 2007, Klanker et al., 2013). Overall, both spike-field coupling and decoupling patterns appear to correlate with potentiation of behavioral outcomes that results in maintenance of choice value signaling or marks unexpected changes in contingencies.

4.3 Conclusion

Consistent with previous studies utilizing lever and spatial rodent approaches, we found that neuronal firing responds to value expectancy in a complex visual reversal task. Spike-firing responses during visual reversal are sensitive to prior reward outcomes, supporting that the OFC is continually monitoring value. Additionally, patterns of behavior were associated with significant alterations in spike- coupling with the local oscillations, which may be a mechanism whereby neuronal responses caused by actions that lead to a desired outcome are propagated to downstream areas required for efficient association learning. Together, this data helps bridge the gap between previous operant and spatial tasks with touch-screen visual learning approaches and provides evidence for spike-field coupling as a putative mechanism of how value expectancy signals may be propagated to influence behavior.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institute on Alcohol Abuse and Alcoholism at the National Institutes of Health (1K22AA020303-01, 1P50AA022534-01 and 5T32AA014127e13).

References

- Amodeo LR, McMurray MS, Roitman JD. 2016 Orbitofrontal cortex reflects changes in response-outcome contingencies during probabilistic reversal learning. *Neuroscience*.
- Bissonette GB, Martins GJ, Franz TM, Harper ES, Schoenbaum G, Powell EM. 2008; Double Dissociation of the Effects of Medial and Orbital Prefrontal Cortical Lesions on Attentional and Affective Shifts in Mice. *Journal of Neuroscience*. 28 :11124–11130. [PubMed: 18971455]
- Boulougouris V, Dalley JW, Robbins TW. 2007; Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat. *Behavioural brain research*. 179 :219–228. [PubMed: 17337305]
- Brigman JL, Daut RA, Wright T, Gunduz-Cinar O, Graybeal C, Davis MI, Jiang Z, Saksida LM, Jinde S, Pease M, Bussey TJ, Lovinger DM, Nakazawa K, Holmes A. 2013; GluN2B in corticostriatal circuits governs choice learning and choice shifting. *Nat Neurosci*. 16 :1101–1110. [PubMed: 23831965]
- Brigman JL, Mathur P, Harvey-White J, Izquierdo A, Saksida LM, Bussey TJ, Fox S, Deneris E, Murphy DL, Holmes A. 2010a; Pharmacological or genetic inactivation of the serotonin transporter improves reversal learning in mice. *Cereb Cortex*. 20 :1955–1963. [PubMed: 20032063]
- Brigman JL, Padukiewicz KE, Sutherland ML, Rothblat LA. 2006; Executive functions in the heterozygous reeler mouse model of schizophrenia. *Behav Neurosci*. 120 :984–988. [PubMed: 16893304]
- Brigman JL, Wright T, Talani G, Prasad-Mulcare S, Jinde S, Seabold GK, Mathur P, Davis MI, Bock R, Gustin RM, Colbran RJ, Alvarez VA, Nakazawa K, Delpire E, Lovinger DM, Holmes A. 2010b; Loss of GluN2B-containing NMDA receptors in CA1 hippocampus and cortex impairs long-term depression, reduces dendritic spine density, and disrupts learning. *J Neurosci*. 30 :4590–4600. [PubMed: 20357110]
- Buzsaki G, Anastassiou CA, Koch C. 2012; The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes. *Nat Rev Neurosci*. 13 :407–420. [PubMed: 22595786]
- Buzsaki G, Draguhn A. 2004; Neuronal oscillations in cortical networks. *Science*. 304 :1926–1929. [PubMed: 15218136]
- Cai X, Padoa-Schioppa C. 2014; Contributions of orbitofrontal and lateral prefrontal cortices to economic choice and the good-to-action transformation. *Neuron*. 81 :1140–1151. [PubMed: 24529981]
- Calderone DJ, Lakatos P, Butler PD, Castellanos FX. 2014; Entrainment of neural oscillations as a modifiable substrate of attention. *Trends Cogn Sci*. 18 :300–309. [PubMed: 24630166]
- Clarke HF, Robbins TW, Roberts AC. 2008; Lesions of the Medial Striatum in Monkeys Produce Perseverative Impairments during Reversal Learning Similar to Those Produced by Lesions of the Orbitofrontal Cortex. *Journal of Neuroscience*. 28 :10972–10982. [PubMed: 18945905]
- Cohen MX. 2014a Analyzing Neural Time Series Data Theory and Practice Preface. *Iss Clin Cogn Neurop*. :Xvii–Xviii.
- Cohen MX. 2014b; Fluctuations in oscillation frequency control spike timing and coordinate neural networks. *The Journal of neuroscience: the official journal of the Society for Neuroscience*. 34 :8988–8998. [PubMed: 24990919]
- Copping NA, Berg EL, Foley GM, Schaffler MD, Onaga BL, Buscher N, Silverman JL, Yang M. 2016 Touchscreen learning deficits and normal social approach behavior in the Shank3B model of Phelan-McDermid Syndrome and autism. *Neuroscience*.
- Costa VD, Tran VL, Turchi J, Averbeck BB. 2015; Reversal learning and dopamine: a bayesian perspective. *J Neurosci*. 35 :2407–2416. [PubMed: 25673835]
- DePoy L, Daut R, Brigman JL, MacPherson K, Crowley N, Gunduz-Cinar O, Pickens CL, Cinar R, Saksida LM, Kunos G, Lovinger DM, Bussey TJ, Camp MC, Holmes A. 2013; Chronic alcohol produces neuroadaptations to prime dorsal striatal learning. *Proc Natl Acad Sci U S A*. 110 :14783–14788. [PubMed: 23959891]
- Fries P, Reynolds JH, Rorie AE, Desimone R. 2001; Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*. 291 :1560–1563. [PubMed: 11222864]

- Graybeal C, Feyder M, Schulman E, Saksida LM, Bussey TJ, Brigman JL, Holmes A. 2011; Paradoxical reversal learning enhancement by stress or prefrontal cortical damage: rescue with BDNF. *Nature Neuroscience*. 14 :1507–1509. [PubMed: 22057192]
- Gregoriou GG, Gotts SJ, Zhou H, Desimone R. 2009; High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science*. 324 :1207–1210. [PubMed: 19478185]
- Hamilton DA, Brigman JL. 2015a Behavioral flexibility in rats and mice: Contributions of distinct frontocortical regions. *Genes, brain, and behavior*.
- Hamilton DA, Brigman JL. 2015b; Behavioral flexibility in rats and mice: contributions of distinct frontocortical regions. *Genes Brain Behav*. 14 :4–21. [PubMed: 25561028]
- Hoover WB, Vertes RP. 2011; Projections of the medial orbital and ventral orbital cortex in the rat. *The Journal of comparative neurology*. 519 :3766–3801. [PubMed: 21800317]
- Hvoslef-Eide M, Nilsson SR, Saksida LM, Bussey TJ. 2016; Cognitive Translation Using the Rodent Touchscreen Testing Approach. *Curr Top Behav Neurosci*. 28 :423–447. [PubMed: 27305921]
- Izquierdo A, Brigman JL, Radke AK, Rudebeck PH, Holmes A. 2016 The neural basis of reversal learning: An updated perspective. *Neuroscience*.
- Jang AI, Costa VD, Rudebeck PH, Chudasama Y, Murray EA, Averbach BB. 2015; The Role of Frontal Cortical and Medial-Temporal Lobe Brain Areas in Learning a Bayesian Prior Belief on Reversals. *J Neurosci*. 35 :11751–11760. [PubMed: 26290251]
- Kennerley SW, Behrens TE, Wallis JD. 2011; Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat Neurosci*. 14 :1581–1589. [PubMed: 22037498]
- Klanker M, Post G, Joosten R, Feenstra M, Denys D. 2013; Deep brain stimulation in the lateral orbitofrontal cortex impairs spatial reversal learning. *Behavioural brain research*. 245 :7–12. [PubMed: 23396148]
- Leach PT, Hayes J, Pride M, Silverman JL, Crawley JN. 2016 Normal Performance of Fmr1 Mice on a Touchscreen Delayed Nonmatching to Position Working Memory Task. *eNeuro*. :3.
- Mar AC, Horner AE, Nilsson SR, Alsio J, Kent BA, Kim CH, Holmes A, Saksida LM, Bussey TJ. 2013; The touchscreen operant platform for assessing executive function in rats and mice. *Nat Protoc*. 8 :1985–2005. [PubMed: 24051960]
- Marquardt K, Sigdel R, Caldwell K, Brigman JL. 2014; Prenatal ethanol exposure impairs executive function in mice into adulthood. *Alcohol Clin Exp Res*. 38 :2962–2968. [PubMed: 25581651]
- McMurray MS, Amodeo LR, Roitman JD. 2016; Consequences of Adolescent Ethanol Consumption on Risk Preference and Orbitofrontal Cortex Encoding of Reward. *Neuropsychopharmacology*. 41 :1366–1375. [PubMed: 26370327]
- Mirenowicz J, Schultz W. 1994; Importance of unpredictability for reward responses in primate dopamine neurons. *J Neurophysiol*. 72 :1024–1027. [PubMed: 7983508]
- Moorman DE, Aston-Jones G. 2014; Orbitofrontal cortical neurons encode expectation-driven initiation of reward-seeking. *J Neurosci*. 34 :10234–10246. [PubMed: 25080585]
- Morris SE, Yee CM, Nuechterlein KH. 2006; Electrophysiological analysis of error monitoring in schizophrenia. *J Abnorm Psychol*. 115 :239–250. [PubMed: 16737389]
- Padoa-Schioppa C. 2007; Orbitofrontal cortex and the computation of economic value. *Ann N Y Acad Sci*. 1121 :232–253. [PubMed: 17698992]
- Padoa-Schioppa C, Assad JA. 2006; Neurons in the orbitofrontal cortex encode economic value. *Nature*. 441 :223–226. [PubMed: 16633341]
- Paxinos, KBJ, Franklin, G. The mouse brain in stereotaxic coordinates. London: Academic Press; 2001.
- Porjesz B, Almasy L, Edenberg HJ, Wang K, Chorlian DB, Foroud T, Goate A, Rice JP, O'Connor SJ, Rohrbaugh J, Kuperman S, Bauer LO, Crowe RR, Schuckit MA, Hesselbrock V, Conneally PM, Tischfield JA, Li TK, Reich T, Begleiter H. 2002; Linkage disequilibrium between the beta frequency of the human EEG and a GABAA receptor gene locus. *Proceedings of the National Academy of Sciences of the United States of America*. 99 :3729–3733. [PubMed: 11891318]
- Riceberg JS, Shapiro ML. 2012; Reward stability determines the contribution of orbitofrontal cortex to adaptive behavior. *J Neurosci*. 32 :16402–16409. [PubMed: 23152622]

- Rich EL, Wallis JD. 2016; Decoding subjective decisions from orbitofrontal cortex. *Nat Neurosci.* 19 :973–980. [PubMed: 27273768]
- Roesch MR, Calu DJ, Schoenbaum G. 2007; Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat Neurosci.* 10 :1615–1624. [PubMed: 18026098]
- Saez A, Rigotti M, Ostojic S, Fusi S, Salzman CD. 2015; Abstract Context Representations in Primate Amygdala and Prefrontal Cortex. *Neuron.* 87 :869–881. [PubMed: 26291167]
- Schilman EA, Uylings HB, Galis-de Graaf Y, Joel D, Groenewegen HJ. 2008; The orbital cortex in rats topographically projects to central parts of the caudate-putamen complex. *Neurosci Lett.* 432 :40–45. [PubMed: 18248891]
- Schoenbaum G, Chiba AA, Gallagher M. 2000; Changes in functional connectivity in orbitofrontal cortex and basolateral amygdala during learning and reversal training. *Journal of Neuroscience.* 20 :5179–5189. [PubMed: 10864975]
- Schoenbaum G, Esber GR. 2010; How do you (estimate you will) like them apples? Integration as a defining trait of orbitofrontal function. *Curr Opin Neurobiol.* 20 :205–211. [PubMed: 20206497]
- Stalnaker TA, Cooch NK, Schoenbaum G. 2015; What the orbitofrontal cortex does not do. *Nat Neurosci.* 18 :620–627. [PubMed: 25919962]
- Sul JH, Kim H, Huh N, Lee D, Jung MW. 2010; Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron.* 66 :449–460. [PubMed: 20471357]
- Takahashi YK, Roesch MR, Stalnaker TA, Haney RZ, Calu DJ, Taylor AR, Burke KA, Schoenbaum G. 2009; The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. *Neuron.* 62 :269–280. [PubMed: 19409271]
- Talpos J, Steckler T. 2013; Touching on translation. *Cell Tissue Res.* 354 :297–308. [PubMed: 23949375]
- Thorpe SJ, Rolls ET, Maddison S. 1983; The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp Brain Res.* 49 :93–115. [PubMed: 6861938]
- van der Meer MA, Johnson A, Schmitzer-Torbert NC, Redish AD. 2010; Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. *Neuron.* 67 :25–32. [PubMed: 20624589]
- van Wingerden M, Vinck M, Lankelma J, Pennartz CM. 2010a; Theta-band phase locking of orbitofrontal neurons during reward expectancy. *J Neurosci.* 30 :7078–7087. [PubMed: 20484650]
- van Wingerden M, Vinck M, Lankelma JV, Pennartz CM. 2010b; Learning-associated gamma-band phase-locking of action-outcome selective neurons in orbitofrontal cortex. *J Neurosci.* 30 :10025–10038. [PubMed: 20668187]
- Vinck M, van Wingerden M, Womelsdorf T, Fries P, Pennartz CM. 2010; The pairwise phase consistency: a bias-free measure of rhythmic neuronal synchronization. *NeuroImage.* 51 :112–122. [PubMed: 20114076]
- Wilson RC, Takahashi YK, Schoenbaum G, Niv Y. 2014; Orbitofrontal cortex as a cognitive map of task space. *Neuron.* 81 :267–279. [PubMed: 24462094]
- Womelsdorf T, Schoffelen JM, Oostenveld R, Singer W, Desimone R, Engel AK, Fries P. 2007; Modulation of neuronal interactions through neuronal synchronization. *Science.* 316 :1609–1612. [PubMed: 17569862]
- Womelsdorf T, Valiante TA, Sahin NT, Miller KJ, Tiesinga P. 2014; Dynamic circuit motifs underlying rhythmic gain control, gating and integration. *Nature neuroscience.* 17 :1031–1039. [PubMed: 25065440]
- Yang M, Lewis FC, Sarvi MS, Foley GM, Crawley JN. 2015; 16p11.2 Deletion mice display cognitive deficits in touchscreen learning and novelty recognition tasks. *Learn Mem.* 22 :622–632. [PubMed: 26572653]
- Young JJ, Shapiro ML. 2009; Double Dissociation and Hierarchical Organization of Strategy Switches and Reversals in the Rat PFC. *Behavioral Neuroscience.* 123 :1028–1035. [PubMed: 19824768]
- Young JJ, Shapiro ML. 2011; Dynamic coding of goal-directed paths by orbital prefrontal cortex. *The Journal of neuroscience: the official journal of the Society for Neuroscience.* 31 :5989–6000. [PubMed: 21508224]

Highlights

- Uses in vivo electrophysiology to examine how the OFC tracks choice values during specific stages of reversal characterized by perseveration, ambiguity and well-learned behavior
- Examines coordination between spike firing and local field oscillations as a mechanism by which choice values are propagated to downstream regions
- Establishes the validity of using touch-screen visual learning paradigms to model OFC-mediated clinical impairments in behavioral flexibility in rodent models

Figure 1. Behavioral Paradigm and Recording Sessions Timeline

A–B. Trials were initiated through a lever press (1) which led to onset of pairwise visual stimuli on a touch-sensitive screen (2). Touch of the rewarded stimulus resulted in delivery of reward in the magazine (3) concomitant with 1 sec tone (3b) and illumination of the magazine light. Touches at the unrewarded led to illumination of the house light (4) for 10 sec. for an incorrect response. Error choices were followed by correction trials in which a subsequent initiation led to the stimuli presented in the same spatial orientation until a correct response was made to prevent side-bias and measure perseveration. **C.** Trials were sorted into four distinct categories, based on the outcome of the previous trial: rewarded trial following rewarded trial= *win-stay*, rewarded trial following error= *lose-shift*, error trial following error trial= *perseverative*, and error trial following rewarded trial= *regressive*. Trial types can also be generally categorized by outcome (rewarded vs. unrewarded) and strategy (stay vs. switch). **D.** After training and electrode implantation mice were recorded at 4 distinct learning stage. Discrimination and Reversal criterion= 85% correct, Early reversal= First session of reversal with performance <20% correct, Chance reversal= ~50% correct. Recording sessions are marked by arrows. **E.** Array image and verification of electrode placement with each point representing the midpoint location of the electrode bundle. Adapted from (Paxinos and Franklin, 2001).

Figure 2. Microstructure of Behavior Matches Learning Stage during Recorded Sessions
A. Cumulative trials required to attain each stage of discrimination and reversal learning. **B.** Total error trials significantly increased during the first session of reversal and subsequently decreased significantly across each subsequent stage of reversal. **C.** *Win-stay* closely reflected total percent correct, significantly decreasing during early reversal and significantly increasing across reversal sessions. **D.** Cumulative total errors to attain each stage of discrimination and reversal learning. **E.** *lose-shift* trials did not significantly increase until chance reversal, before significantly decreasing again at criterion. **F.** *Regressive* error trials, another switch trial type, closely mirrored *lose-shift*, increasing at chance before decreasing again at criterion. Data are means $\pm$ SEM. N=12 mice per recording session. * = $P < .05$, compared to Discrimination criterion; # = $P < .05$, compared to early reversal; and † = $P < .05$, compared to chance reversal.

Figure 3. Average Spike-Firing Rate and Phasic Neuron Recruitment for each Trial Type is Specific to Learning Session

A. Discrimination Criterion: Average firing significantly increased during the 2–3 sec epoch for both reward trial responses, with *win-stay* firing increasing significantly more than *lose-shift* (left). Average neuronal firing significantly increased across all post-choice epochs for both error trial types (center). Significantly more neurons were responsive to *win-stay* versus *lose-shift* rewarded trials, while there was no difference between number or error responsive neurons (right). **B. Early Reversal:** There was no significant increase in average firing for either rewarded trial type for any epoch (right). Both error trial types significantly increased across all post-choice epochs (center). Significantly more *lose-shift* responsive neurons were seen versus *win-stay* and significantly more *perseverative* than *regressive* error responsive neurons (right). **C. Chance Reversal:** Both reward trial types had increased firing to cue onset but neither was significantly above baseline (left). Neither error trial type showed significant increases in average firing (center). No significant difference was seen between event-responsive neurons to reward trials while *regressive* responsive neurons were significantly greater than *perseverative* errors (right). **D. Reversal Criterion:** Average firing significantly increased during the 2–3 sec epoch for both rewarded trial responses with firing after *win-stay* increasing significantly more than *lose-shift* (left). No significant increase was seen in average firing rate for error trials (center). Significantly more *win-stay* responsive neurons were seen with no *lose-shift* responsive neurons recorded and no difference in percentage of error responsive neurons (right). X-axis indicates if the responsive neuron was found for a correct response (rewarded cue) or incorrect response (error cue). Data are means $\pm$ SEM. Gray shading indicates the presence of the associative cue: tone (correct) or

house-light (error-cue). Bar indicates significant difference from baseline by type (color). *
= $P < .05$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4. Spike- Oscillatory Synchrony Pattern is Distinct During the Post-Choice Phase of Each Learning Session

A. Discrimination Criterion. *Lose-shift* trials showed significant increases in spike-field coupling in the theta frequency while *win-stay* trials showed significant decoupling in the beta frequency (upper). *Regressive* error trials had significant increases in the theta range while *perseverative* responses showed significant decoupling across all frequencies (lower). **B. Early Reversal:** *Lose-shift* trials showed no significant coupling or decoupling across frequencies while *win-stay* trials now showed decoupling across Theta, the majority of Beta and all Gama frequencies (upper). Neither error trial type showed significant coupling or decoupling during this stage (lower). **C. Chance Reversal:** *Win-stay* trials showed no significant coupling or decoupling while *lose-shift* now had strong coupling in the Theta frequency and significant decoupling in the mid Beta range (upper). While more variable, neither error trial type showed significant coupling or decoupling (lower). **D. Criterion Reversal:** While increasing from chance, *win-stay* trials showed no significant coupling or decoupling while *lose-shift* trials now showed significant decoupling in all frequencies (upper). *Regressive* errors now had significant decoupling across Beta and Gama frequencies, while non-measurable coupling response was seen to *perseverative* trials (lower). Dotted Lines mark boundaries of frequency ranges of theta (θ , 4–10 Hz), beta (β , 10–30 Hz) and gamma (γ , 30–40 Hz). Data are means $\pm$ SEM. Colored lines indicate the frequency ranges of significance for the corresponding color trial types. *= significant difference ($P < 0.05$) from chance as calculated by two standard deviations away from random shuffled distribution of the data (grey bar).